DOCUMENT RESUME

ED 031 396 SE 006 658

By-Welch, Wayne W.

Applied Research in a Curriculum Project: Accomplishments and Limitations.

Harvard Univ., Cambridge, Mass. Harvard Project Physics.
Pub Date Feb 69
Note-10p.; Paper presented at the meeting of the American Educational Research Association, Los Angeles, California, February 1969.
EDRS Price MF-$0.25 HC-$0.60
Descriptors-*Curriculum Development, *Evaluation, *Physics, *Science Course Improvement Project, Science Education, *Secondary School Science
Identifiers-Harvard Project Physics

The validity of statements by various "national curriculum projects" concerning the availability of research evidence to support their claims for the attainment of their objectives is questioned. Evaluation procedures employed by these various curriculum development groups are criticized for lack of empirical data. Obtaining such data is seen as the purpose of applied research in curriculum projects. Applied research differs from basic research in that it is oriented to a specific curriculum, rather than to variables common to many curricula. Two components of applied research are "formative evaluation" and "summative evaluation," the former concerned with questions concerning the product while being developed and the latter concerned with the final product. The success and limitations of both formative and summative evaluation techniques used by Harvard Project Physics (HPP) is described. In process is a book for the professional researcher describing in detail HPP procedure and the results of both formative and summative evaluation studies. (BC)

APPLIED RESEARCH IN A CURRICULUM PROJECT: Accomplishments and Limitations

Wayne W. Welch
Harvard Project Physics

There is general agreement that an important part of national curriculum projects is a comprehensive program of evaluation. Although many projects give lip service to this goal, there is little evidence to support their claims. Of 68 projects reviewed in the 1968 Report on Science and Mathematics Curricular Developments,[1] only 19 indicate the availability of research evidence demonstrating achievement of their objectives. But since only 6 of these 19 bothered to use randomly selected control groups, one seriously questions the generalizability of the results that are available.

Evaluation as conducted by national projects for the most part has consisted of large quantities of feedback and some achievement testing. The feedback is usually unstructured and typically includes teacher reactions, classroom visits by staff, anecdotal reports and professional opinion. All 47 of the 68 projects who reported conducting any kind of evaluation included one of these activities. While feedback information may give project personnel some feeling of whether or not they are achieving their hopes and ambitions, it is fraught with dangers of sentiment and subjectivity. This unstructured feedback, so often the only evaluation conducted by a curricular development group, reminds me of playing a basketball game, not keeping score and then circulating among the spectators asking opinions on who won the game.

Paper presented at the 1969 annual meeting of the American Educational Research Association, Los Angeles, California, February 6, 1969.

On second thought, because control groups are seldom used, a better analogy might be only one team playing, and then asking spectators what they thought about the teams' chances for success. The point to be made out of these analogies is that evaluation opinion needs to be supported by some hard-nosed data. Obtaining this data is the purpose of applied research in curriculum projects.

"Applied research" includes those activities that are designed to gather information useful in making decisions about a specific curriculum or course. Applied research differs from basic research in being oriented to a specific curriculum, rather than to variables common to many curricula. Basic research might involve the question "What, if any, is the effect of teacher personality on the social climate of learning?" This is a general question, appropriate to many courses. Applied research, on the other hand, invclves questions about specific programs: for example, "What mathematical knowledge is prerequisite to success in the CHEM study course?" The latter question is of interest primarily to those developing the course and to those who are considering using it.

On many occasions, applied and basic research interact with and complement each other. But for the purposes of this paper, they will be treated as separate endeavors. It is useful to think of two separate components of applied research. One is designed to answer questions about improving a curriculum during its development (sometimes called "formative evaluation"). The other is designed to answer questions about the final product (usually called "summative evaluation").

Because we have spent a great deal of time and energy on both these aspects of evaluation, it may be of some interest to outline some of the advantages and pitfalls we have identified.

Formative evaluation is conducted for the purpose of course improvement. Most projects usually do some kind of casual formative evaluation if provisions for revision are written into their development plans. In addition to the usual verbal and highly subjective kinds of formative evaluation, it is possible to carry out more structured and objective investigations. For example, studies of readability, the length of materials, achievement test results, semantic differential type questionnaires, inventories, student questionnaires, ratings of teachability and interaction analysis all can be used to some extent. But each of these methods has its own strengths and limitations and the evaluator must be careful not to fall into the trap of believing his program is complete just because it is based on one or several of these methods. Information gathered from each kind of evaluation study can be used to answer different questions.

To give some idea of the success and limitations of a particular type of formative evaluation study, I have chosen one of the several questions investigated by our evaluation group and also carried out by several other projects. Namely, "Do the results of the achievement tests indicate areas where the course could be improved?" I want to describe our procedure briefly and point out some of the problems and successes we encountered.

In the second trial year of the Project, an achievement test was written for each of the six units of the course. To provide information on which to base revisions, the authors were given item analyses of each of the tests. In this way, various parts of the text that were not conveying their message to the students would be identified and authors could concentrate their efforts on improving these areas.

The test writing process had several practical problems to overcome, perhaps most important was the severe time restriction on producing the tests. Only about six weeks were available to write, revise, print, and have each test to the teacher by the time the unit was finished.

The course was being tried by 16 teachers and 500 students from all parts of the country so test arrival and responses were subject to the whims of the U.S. Post Office. However, we found that it was possible to write a test, have it printed, mailed, and in the teacher's hands within the six week limitation.

Data were obtained on each of the tests (subject to the varying promptness of teachers returning the answer sheets), and a complete item analysis performed. Means, standard deviations, item discrimination indices, success levels, and test reliabilities were computed. Item success levels were sent to revision authors together with the percentage of students choosing each distractor. Thus, we had established what I consider a typical example of formative evaluation, and one that I thought would be found in several curriculum projects.

However, only 7 of 68 projects reviewed in the AAAS Report indicated they had used test results in this manner.

This procedure enjoyed some limited success. Several items with high difficulty levels indicated sections of the text that required rewriting. A few items helped identify topics where existing explanation seemed adequate. However, there were practical problems in this sequence that future curriculum groups should be aware of.

Perhaps the greatest problem is the time it takes to receive answer sheets from teachers scattered throughout the country, key punch the results, and run the cards through the item analysis program. Regardless of our efforts to shorten it, approximately three or four months were required. Unfortunately, a large portion of the text revision would already be completed by the time the item analysis results were available to the authors. Furthermore, several authors were post hoc suspicious of the test items and their ability to probe the understandings the authors believed to be in the text. If a set of item statistics indicated a section was not being understood, these authors were more likely to criticize the tests, than to criticize the text. This was sometimes the case even though they initially had approved the items.

Another problem we encounterd is the extreme difficulty an author has in trying to make a concept or idea more understandable once it has been determined that students have failed to grasp the idea. He may know that students do not understand Newton's Third Law of Motion, but may not be able to revise the text to make it simpler. It is

entirely possible the revision will make the topic even more difficult.

These comments on the value of this type of formative evaluation do not imply failure, but rather the existence of limitations. I believe that it does help improve a course and has a definite place in an evaluation program. However, we must recognize that its usefulness is limited by a number of practical factors, and that these factors must be considered when determining the relative emphasis, timing, and importance to place on formative evaluation.

I would now like to turn to the summative aspects of applied research in curriculum projects. These are activities carried out for the purpose of providing information to the eventual user of the course. This information is valuable to the user (teacher, supervisor, etc.) in so far as it helps him make decisions concerning adoption and effective use of the course. In my opinion all projects have an obligation to describe their programs and to provide evidence of success in achieving stated or implied objectives.

Examples of several summative evaluation questions studied in connection with Project Physics include:

1. You state that one of the goals of your course is to increase enrollments in high school physics. What evidence do you have that this objective is being achieved?

2. In order to increase enrollments in physics, you must appeal to groups of students who normally do not take physics. How do these students perform in your course?

3. What teacher preparation is required to teach this new program effectively?

4. How do Project Physics students perform on national examinations such as the College Boards?

5. Has the course been trial tested using typical physics teachers, rather than using selected physics teachers?

6. What changes in attitude and interest have been identified as a result of taking your course?

To illustrate an example of summative evaluation and to point out its success and limitations, let me describe a study we conducted to answer the question, "What appeal does the course have to those students now successfully avoiding any study of physics?" Of special interest are girls since 95 out of 100 usually choose not to take physics. We wanted to determine if they responded more favorably to Project Physics than to other courses.

During the 1966-67 trial year, a pilot study examined the relative success of boys and girls. A similar study was conducted again during 1967-68 using a ·tional random sample of teachers with random assignment to experimental and treatment groups. Included in the 36 schools that were using the Project Physics materials for the first time were 322 girls. In the 21 comparison schools there were 159 girls. In both groups the number of girls is about 24% of the total physics population. Also, the initial interest in physics and IQ of the experimental and control groups were approximately the same.

Without going into detail regarding the instruments and method of analysis, we found that, indeed, girls in Project Physics did seem to respond more favorably to that course. For example, on a measure of course satisfaction, they had a mean rating of .60 while the comparison group rating was .47. As expected, Project Physics girls gained more on our achievement test, but they also showed greater

gains on several non-project cognitive measures. They saw physics as more diverse and less difficult than did the comparison group. And they received from their physics teachers higher course grades, certainly an important factor in determining course satisfaction.

We examined change in interest in the subject over the year. We found that both groups of girls showed declines in interest as measured by several different tests. For example, on a semantic differential composite interest score, Project Physics girls declined .5 pre test standard deviations while the comparison group declined .6 standard deviations. Loss in interest is certainly not what course developers have in mind; it is mildly compensating that Project Physics girls did not decline as much as the comparison group. I should mention that several other studies have noted similar declines in interest in school subjects.[2,3] Reporting this finding to prospective Project Physics teachers might be hazardous, but we are hoping a straightforward presentation of all findings will not only indicate an honest evaluation program, but also help to increase our understanding of physics education.

There are limitations in summative evaluation in addition to the possibility of negative findings. Of course we are limited in our evaluation by the precision of our instruments. There is some indication of ceiling effects on the semantic differential test. We have also encountered problems of attrition; a teacher who was supposedly using our program was in fact using a traditional textbook, tests were lost during mailing, on a few occasions teachers did not

give tests, and perhaps most important, there were some serious delays on the part of our production staff to provide teachers with the necessary materials in time to use them in the schools. These practical problems are very much a part of evaluation and impose their own peculiar limitation -- certainly not insurmountable, but always in evidence to the applied researcher.

It is too early to assess the impact of our summative evaluation findings. We are writing a book for the professional researcher describing in detail our procedure and the results of both formative and summative evaluation studies. A shorter pamphlet written for teachers, administrators, and guidance counsellors will also be available to those inquiring about the course. We hope the information they find there will help them to make rational decisions regarding adoption and use of the course. Once these documents are available and tried by teachers we will be in a better position to assess the accomplishments and limitations of applied research in our curriculum project.

References

[1]Lockard, J. David, Sixth Report of the International Clearinghouse on Science and Mathematics Curricular Developments 1968,(A joint project of the American Association for the Advancement of Science and the Science Teaching Center, University of Maryland).

[2]Flanders, Ned A., Morrison, Betty M., and Brode, E. Leland, "Changes in Pupil Attitudes During the School Year," Journal of Educational Psychology, Vol. 50, No. 5, 1968, p. 334-338.

[3]Atkin, J. Myron and Wyatt, Stanley P. Jr., "Elementary-School Science Project" (final report), (University of Illinois, Urbana, June 1968), p. 57ff.